A Polar-based Logo Representation based on Topological and Colour Features

Farshad Nourbakhsh
Computer Vision Center
Edifici O - Campus Universitat
Autonoma de Barcelona (UAB)
Bellaterra, Spain
farshad@cvc.uab.es

Dimosthenis Karatzas
Computer Vision Center
Edifici O - Campus Universitat
Autonoma de Barcelona (UAB)
Bellaterra, Spain
dimos@cvc.uab.es

Ernest Valveny
Computer Vision Center
Edifici O - Campus Universitat
Autonoma de Barcelona (UAB)
Bellaterra, Spain
ernest@cvc.uab.es

ABSTRACT

In this paper, we propose a novel rotation and scale invariant method for colour logo retrieval and classification, which involves performing a simple colour segmentation and subsequently describing each of the resultant colour components based on a set of topological and colour features. A polar representation is used to represent the logo and the subsequent logo matching is based on Cyclic Dynamic Time Warping (CDTW). We also show how combining information about the global distribution of the logo components and their local neighbourhood using the Delaunay triangulation allows to improve the results. All experiments are performed on a dataset of 2500 instances of 100 colour logo images in different rotations and scales.

Keywords

Colour logo retrieval and logo representation

1. INTRODUCTION

A logo is a graphical element designed for easy and definitive recognition, and is typically used to identify a company or organisation [18]. Logos can be found in many kinds of paper documents, scene images, web images, signs and banners, and their recognition and extraction are demanding tasks in computer vision and document analysis. Although logos can be found in many styles but they are bound by certain design restrictions as they need to be salient and easily identified. They are usually rendered in colour and may, or may not, contain any textual content. Depending on their use they might appear in different scales, rotations or 3D settings. Developing robust and accurate algorithms for the location and recognition of logos in documents would present many advantages including improved document retrieval and classification, better compression techniques and superior mass document manipulation and management. As such, logo recognition has received a lot of attention during the past several years and it remains a hot research topic.

DAS '10, June 9-11, 2010, Boston, MA, USA

Numerous approaches to shape analysis, with emphasis on logo recognition, have been reported in the domain of grey or bi-level document image analysis [19], [8], [22], [7], [2], [11], [3], [9], [5], [14] and [6]. For example, Zhu and Doermann [20] propose a method for logo detection in document images that combines contextual information about the expected logo location in the document with geometrical features of extracted components (aspect ratio, spatial density and area). They use a Fisher classifier at different scales of the image and calculate performance by comparing the overlapping of the resulted logo regions with ground truth ones. They report high precision and recall values for the Tobacco-800 dataset and highlight the importance of multiscale approaches.

In [21], Zhu and Doermann report a new method, that makes use of the localisation capabilities of [20] in coarse image scales to boost a cascade of classifiers at finer image scales. They compare different algorithmic variants based on different shape matching techniques (shape context and neighbourhood graph matching) combined with different shape dissimilarity measures (including two new measures based on the anisotropic scaling between two shapes), and report similar results as in [20] on the same dataset.

Nevertheless, although shape is undoubtedly significant, colour is an equally important cue for logo recognition which, more often than not, is not taken into account. The few methods that exist on colour logo location and recognition are generally treating real scene images, where the types of deformations (e.g. perspective distortion, occlusions etc) are different from what is expected in scanned documents and digitally born images. Although the authors report results on datasets of colour images, in many of these methods, the images are first converted into grayscale and the methods really work on the grey information.

For example, Phan et al. [17], [15] proposed a logo and trademark detection method in unconstrained colour images based on edge gradient information. They calculate the colour edge gradient co-occurrence histogram (CEGCH), an extension of the colour edge co-occurrence histogram (CECH [13]), by using vector order statistics. To accomplish this, the query image is quantized by a HSV colour quantization method and the information is extracted by CEGCH. In the next step, multiple scales of the corpus images are considered and the above procedure is applied to each of them. Finally, overlapping search windows are used to search for the input query in the dataset images. The CEGCHs of every search window at each scale factor are computed and compared to the input query CEGCH to localise the logo.

In [16] further experiments and more comprehensive comparision is presented. The authors demonstrate that an improvement can be achieved by using CEGCH over the use of CECH. They also provide a critical appraisal of their method, highlighting the areas where further improvements are needed, namely the inability of the algorigthm to locate multiple logos in the same image and the limitations in the degree of logo deformation that can be tackled.

A. Hesson and D. Androutsos [12] proposed a method which is based on [17], although this time they apply the Haar transform on the grey scale version of the quantized down sampled image, and use the wavelet decomposition coefficients as means of capturing edge information. The do limited evaluation, using two query images on a dataset of 2000 pre-segmented logos and report very low retrival rates.

Z. Ahmed and H. Fella [1] have proposed a method for finding potential logo locations in document images based on shape and chromatic properties of logos. They first threshold the image and locate compact black areas in the bi-level version, and subsequently examine the colour content of each extracted areas under the assumption that logos comprise a few uniformed coloured regions. A number of assumptions about the logo images are inherently made to construct the search heuristics which make the method sensitive to noise.

In this paper we present a rotation and scale invariant method for colour logo retrieval and classification. The method is based on an initial colour segmentation of the logo into connected components which are subsequently described using a set of geometric, topological and colour features. The set of resulting components are used to construct a polar logo representation. The Cyclic Dynamic Time Warping (CDTW) is then used for comparing logos, as the polar representations have variable lengths. Two flavours of logo representation are introduced that take into account topological information in different ways. The different representations are tested on a synthetic dataset that comprises 2500 instances of 100 colour logos in different rotations and scales.

This paper is organized as follows: In section 2 we explain the preprocessing steps of colour segmentation and filtering by which we obtain the connected components utilised in the logo description. In section 3 we detail the different logo descriptors while section 4 introduces the logo matching process. In section 5 the different descriptors are compared and results are presented in a retrieval scenario. Finally, section 6 concludes the paper with a critical appraisal of the obtained results.

2. IMAGE PRE-PROCESSING

Logos generally comprise a small number of uniform coloured regions. These regions carry adequate information to properly describe a logo image, hence we decided to use the connected components resulted from a colour segmentation as the building blocks of our logo representation. In order to extract the most representative logo regions, an initial set of connected components is produced through colour segmentation and the resulting components are subsequently filtered to eliminate non-relevant ones from the description. These steps are described below.

2.1 Colour Segmentation

The fact that logo images generally comprise uniform coloured regions is an advantage as it ensures a mostly error-free segmentation. Hence a relatively simple colour segmentation method has been employed here, which produces reasonably stable results.

The segmentation process used here is a one-pass algorithm which creates 8-connected components based on colour similarity. The algorithm processes the image in a left-to-right, top-to-bottom fashion, and for each pixel calculates its colour similarity with the four of its neighbours that are already assigned to a connected component. The pixel then is either assigned to one of the existing components if their colour difference is below a set threshold, or it is used as the seed for a new component. Additional checks are performed in case a pixel is similar to more than one existing component, in which case components might be merged. The algorithm is further described in [4].

Colour similarity is assessed in the RGB colour space. Since uniform coloured regions are sought working with RGB results in a considerable faster algorithm without seriously affecting performance. After several tests with the images of our dataset, we set the colour threshold to 130 as it permits to get components that are robust to rotation, scale and illumination.

2.2 Component Filtering

After segmenting the image into a set of connected components. We apply a filtering step with the aim of eliminating small components which do not carry useful information for logo description and can hinder the matching process. Another type of noise that can appear is due to anti-aliasing artefacts in the original image. Typically anti-aliasing will result into components with a width of 1-2 pixels. These components do not carry any useful information, therefore they are discarded.

2.3 Background Elimination

The next step after component filtering is to extract the foreground components. Hence, the components that touch the border of the logo image are selected as potential background components. If the number of touching pixels in these selected components is larger than half of the height or the width of the logo, the colour of the selected component is picked as a background colour. Following that, the colour of each connected component is compared against the background colours and if similar is also labelled as background and is removed. Figures 1 and 2 show one of the logo images in the dataset and its result after colour segmentation where each component is illustrated in a different colour. Figures 3 and 4 depict the result of component filtering and background elimination respectively.

Figure 1: The original image

3. LOGO REPRESENTATION

After obtaining foreground connected components, we extract a set of features for each of the components and define a representation which can convey the topological information between these components. In detail, we first extract a set of features for each connected component, then investigate different ways to represent the logo as a whole, based on the individual components. Then, we have used a polar-based representation where all components are linked to a given reference point and sorted clockwise according to the angle to this reference point (Figure 5 and Figure 6). Given this representation, the key points are first how to select the reference point (section 3.1) and then whcih set of features to use to represent each connected component (section 3.2). In addition, we have also investigated an extension of this basic representation adding information about the local neighbourhood of each connected component using the Delaunay triangulation (section 3.3)

3.1 Reference Point

Two possibilities for the logo reference point are to use the centre of the biggest connected component or the center of mass defined by all the components of the logo.

1. Biggest Connected Component: The center of mass of the biggest connected component, the foreground component with the highest number of pixels, is selected as the reference point. The main advantage of choosing the biggest connected component centre as a reference point is that this component is more robust to different variations, so it is easier to locate it under different scales and rotations.

2. Center of Mass: Choosing a reference point that depends on the existence of a specific component in the segmentation results can be risky even if it is the most robust one. Earlier work on document classification [10], where a polar representation is used, suggested the usefulness of center of mass of the combination of all foreground components as a good reference point. The advantage is that as it is computed over all the foreground components it could be a more stable choice.

3.2 Connected Component Representation

Each connected component in the logo description is represented by a feature vector. A number of features were examined that express both the connected component's own

Figure 2: After color segmentation

Figure 3: After the component filtering

Figure 4: Separating the biggest background component

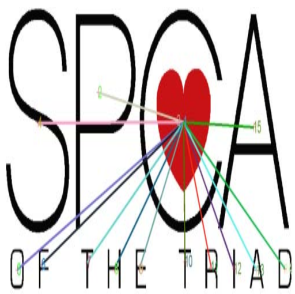

Figure 5: Biggest connected component representation

Figure 6: Center of mass representation

properties (i.e. its size and colour) as well as the spatial relationships between connected components. The features assessed are listed below:

1. Forward Angle: This is the angle difference between two consecutive connected components ordered in clockwise format and the reference point. This vector is normalized to one by dividing by 360 in each logo.
2. Backward Angle: This feature is like the forward angle but this is the angle with respect to the previous connected component. Like the forward angle, it is normalized to one.
3. Distance between every component and the reference point: This feature vector is normalized by dividing by the maximum distance value.
4. Average colour value of each connected component: Each connected component has a RGB colour value which is the average RGB colour value of all pixels in that component. It makes no sense to normalise the colour components individually, so normalisation in this case takes place at the time of calculating the colour distance between two points (see section 4.1). For normalization, the colour distances will be divided by the maximum possible colour distance of $255\sqrt{3}$.
5. Total Number of pixels in each connected component: This feature is normalized by dividing by size of the largest component in each logo image.

3.3 Local Neighbourhood Representation

The previous representation provides global (logo level) information about the distribution of the components in the logo without paying attention to the local (component level) relationships among them. On the other hand, the global information might not be robust enough to possible distortions. Thus, we attempted to construct a more robust representation by adding some local information about the neighbourhood of each connected component to the logo representation. In this sense, we extended the basic representation described above devising a two-level representation based on the Delaunay triangulation of the centers of the foreground connected components. The Delaunay representation permits to create a unique geometry of the local neighbourhood of the connected components. Then, each connected component can be taken as a new local reference point and all its neighbouring components according to the Delaunay triangulation can be sorted and represented using the same polar-based representation as used before. Subsequebntly, the two representation levels, global and local, are combined in the following way as shown in Figures 7-11:

Figure 7: The original image

- After applying the colour segmentation and background separation, the center of each foreground connected component is specified (Figure 8).
- The components are ordered based on the chosen logo reference point in a polar fashion (center of mass or biggest connected component) (Figure 9).
- The Delaunay triangulation is applied to these center points (Figure 10).
- For each connected component, its Delaunay triangulation neighbouring are ordered in a polar fashion (Figure 11).
- All the features presented in the previous section are used as connected component descriptors.

Figures 7-11 illustrate the above steps with more details. Figure 9 shows the center of each connected component as ordered in respect to the logo reference point (in this case, the center of mass is used). Figure 11 shows the new ordering at the local level based on the polar format. In more detail, the neighbours of component #3 are marked by different numbers based on the Delaunay triangulation. These neighbours are ordered based on the polar format before feature selection.

4. LOGO MATCHING

Having obtained a logo representation, the next step is logo matching. The main problem here is that the length of the description is not the same for different logos. A possible approach for calculating similarity between unequal sequences is Dynamic Time Warping (DTW) which defines a dissimilarity measure based on an optimal alignment of two (non-cyclic) strings and has been successfully applied to speech recognition, on-line handwritten text recognition and

Figure 8: After color segmentation with specified center point of each foreground components

Figure 9: Ordering them based on the chosen reference point

Figure 10: After applying Delaunay triangulation

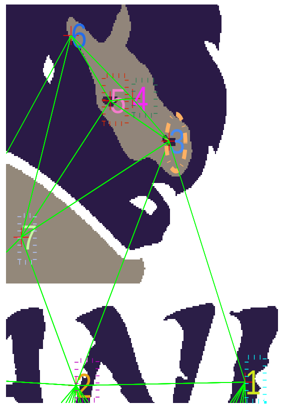

Figure 11: Shows one connected component and its neighbours respect in Delaunay triangulation

time series alignment. For example, Gordo and Valveny [10] have presented a rotation invariant page layout descriptor based on cyclic dynamic time warping (CDTW). This work is a good example of using DTW for comparing descriptions which do not have the same length.

4.1 One level matching

In order to apply the DTW, we need the definition of a function cost between every pair of connected components (points in the DTW sequence). This function cost will be different for each of the two representations explained above. In the basic representation, to calculate the difference between connected components (points in the DTW sequences) we are using a weighted sum of the difference between each component feature.

$$\begin{aligned} D(a,b) &= W_1 \cdot d(f_{a_1}, f_{b_1}) + W_2 \cdot d(f_{a_2}, f_{b_2}) + \\ &+ \dots + W_n \cdot d(f_{a_n}, f_{b_n}) \end{aligned}$$

where $W_i = 1/n$ for all $i \in \{1, \dots, n\}$. In the above equation f_{a_i} is each of the component features described in section 3.2, W is the weight and $D(a,b)$ is the distance between components a and b that is used in CDTW.

4.2 Two Level Matching

In case of the extended representation, based on the Delaunay triangulation, the distance between two logos is obtained after applying twice the CDTW, one at the local level of each connected component and its local neighbours and another one at the global level of the logo. At the local component level, the CDTW is applied using the same cost function described in the previous section. Then, this distance is used as the cost function between two components in the CDTW applied at the logo level.

5. EVALUATION

To validate the system, we first describe the data and experiments for two application scenarios: recognition and retrieval. First, we will give some details about the dataset we have used (section 5.1). Then, we will use the basic representation (without local information as provided by the Delaunay triangulation) with the biggest connected component as a reference point to determine the best set of features to represent each connected component (section 5.2) and to evaluate the effect of the number of connected components in the logo in the retrieval performance (section 5.3). Finally, we will provide a comparison of the four possible logo representation (basic vs extended and biggest component vs centre of mass reference point) according to their retrieval performance (section 5.4).

5.1 Data Set

Because the amount of work on logo recognition in the colour domain is very limited colour logo datasets are sparse. Therefore, we decided to create our own data set. To do so we used Google picture search with the key word "Colour Logo". The first 100 logos returned were included the dataset. Our dataset is composed by different logo images resolution which are between (215×54) and (578×697). Additionally, we were not generally able to find different rotations and scales of the chosen logos on the web. Hence we created different rotation and scale settings for each logo in our dataset:

The image of each class was scaled by the factors of 65%, 150%, 135%, 125% and 85% and each resulting image rotated at 9, 36, 45, 90 and 150 degrees resulting in 2500 instances of the original 100 logos.

5.2 Evaluation of Different Feature Sets

For this experiment, the biggest connected component based logo representation is assessed on the whole dataset (2500 instances) to investigate the performance in a classification scenario when employing different feature sets and their combinations. Specifically, we have different combinations of features to describe connected components (CC). When combined, the individual features are considered with equal weights. Table 1 shows the result of possible combinations. It is clear, after using all 5 features the accuracy increased from individual features to 81.44%.

Feature Set	Accuracy
The forward angle	41.2%
The backward angle	46.6%
The distance of each CC from the RP	47.84%
The number of pixels of the components	32.76%
The colour of the components	51.88%
The combination of above	81.44%

Table 1: Accuracy obtained with respect to the different feature sets for 2500 instances.

5.3 Effect of Logo Complexity

As in the previous section, we have used the same dataset and the basic representation with the biggest connected component reference point, only now in the retrieval scenario. The aim of this experiment is to investigate the performance in relation to the complexity (number of connected components) of a logo. We expect that the more components available in a logo, the richer the logo description is so the result of logo matching would be better. The 100 original logo images are sorted based on the number of their connected components. The maximum one was 210 and the minimum one was 2. Then the set of images is divided into 3 classes based on the number of connected components (each class contains a different number of images), Table 2 shows the number of logos, connected components and the value of the Mean Average Precision (MAP) metric for each class based on the Precision and Recall definition.

The precision and recall are defined based the following equations.

$$P = \frac{|ret \bigcap rel|}{|ret|} \tag{1}$$

$$R = \frac{|ret \bigcap rel|}{|rel|} \tag{2}$$

Where $|ret|$ is the number of returned logos and $|rel|$ is the number of relevant logos. In addition the precision based on recall are called precision-recall plot which is used to plot figures 12 and 13.

Mean Average Precision (MAP) rewards ranking relevant earlier. It is defined as the average of precisions computed at the point of each of the relevant results in the ranked sequence and given by below equation:

$$MAP \quad = \frac{\sum_{r=1}^{|ret|} (P(r) \times rel(r))}{|rel|} \tag{3}$$

Where r is the rank, rel is a binary function on the relevance of a given logo and P is the precision at a given cut-off rank. The distance between the query logo and the test images, as calculated in section 4, is used to rank the results. Figure 12 depicts the dependency of the proposed method on the number of connected components. In detail when the number of components is lower than a certain amount, we do not have enough information to describe the logo as Figure 12 shows. When the number of connected components is small the retrieval results are worse.

Classes	Number of CC	Number of Logos	MAP
A	41-210	20	0.9461
B	21-40	31	0.9514
C	2-20	49	0.8839

Table 2: Shows the number of logo and connected components in each class

5.4 Comparison of Different Logo Representations

The same rotation-scaled dataset (2500 instances) and precision and recall strategy are used to investigate the performance of our method in the retrieval scenario for all four possible logo representations (using the center of mass or the biggest connected component as a reference point and using the basic representation or the extended representation with the local information obtained from the Delaunay triangulation). Figure 13 shows the precision vs recall plot for the above experiments and gives a comparison between

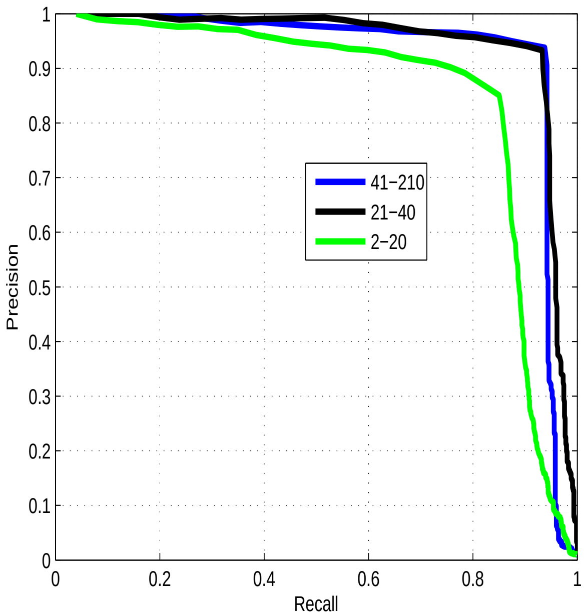

Figure 12: the resultant Precision-Recall plot of each divided class

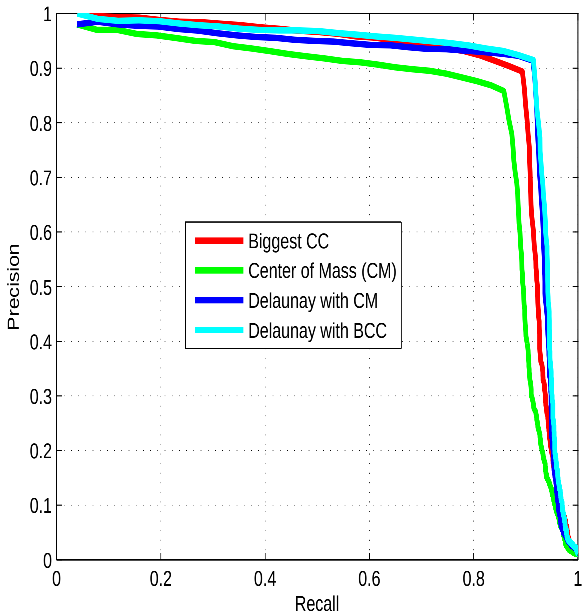

Figure 13: Precision-recall result curved for biggest connected component, center of mass and Delaunay triangulation method

the explained representations. The result of the representation based on the Delaunay triangulation using the biggest connected component as a reference point is better than the other ones based on the value of the mean average precision.

This result is confirmed by the values of the Mean Average Precision (MAP) measure given in Table 3. The reason for this behaviour can be two-fold: on one hand, the biggest connected component yields a more robust reference point. On the other hand, the local information provided by the Delaunay triangulation seems to improve the overall representation. Nevertheless, the two-level representation is computationally slower than using only the global representation.

Representations	MAP
Biggest Connected Component	0.9173
Center of Mass	0.8817
Delaunay triangulation based on the CM	0.9269
Delaunay triangulation based on the BCC	0.9318

Table 3: Shows the Mean Average Precision for precision vs recall plots

6. CONCLUSIONS

In this paper we developed and tested a novel method for colour logo retrieval and classification. The method involves first performing a simple colour segmentation and then describing each of the resultant colour components based on a set of topological and colour features. A polar representation is used to represent the logo and logo matching is subsequently performed, using CDTW. Two flavours of the method are proposed, differing in the selection of the reference point, and a comparison between them is made. We concluded that combining the proposed features gives a better result than using individual features alone. An accuracy of 81.44% was achieved with the combination of 5 features. It is worth to be mentioned that it is difficult to compare our proposed method with the existing methods because these methods are applied for non-color logo datasets. We observed a dependency of the method on the number of connected components in the logos. The optimal result can be obtained when the number of connected components is not too high or too low. The logo representation based on the Delaunay triangulation with the biggest connected component as a reference point gives a better result compared to the other proposed methods. This difference occurs because the Delaunay triangulation representation combines local and global information.

7. AKNOWLEDGEMENTS

This work was partially supported by the Spanish projects TIN2008-04998, TIN2009-14633-C03-03, CONSOLIDER INGENIO 2010 (CSD2007-00018) and the fellowship RYC-2009-05031.

8. REFERENCES

[1] Z. Ahmed and H. Fella. Logos extraction on picture documents using shape and colour density. *IEEE International Symposium on Industrial Electronics, ISIE*, pages 2492–2496, 2008.

[2] F. Cesarini, E. Francesconi, M. Gori, S. Marinai, J. Q. Sheng, and G. Soda. A neural-based architecture for spot-noisy logo recognition. *4th International*

Conference on Document Analysis and Recognition, pages 175–179, 1997.

[3] J. Chen, M. K. Leung, and Y. Gao. Noisy logo recognition using line segment hausdorff distance. *Pattern recognition*, pages 943–955, 2003.

[4] A. Clavelli and D. Karatzas. Text segmentation in colour posters from the spanish civil war era. *In Proceedings of the Tenth International Conference on Document Analysis and Recognition, ICDAR09*, pages 181–185, 2009.

[5] R. den Hollander and A. Hanjalic. Logo recognition in video stills by string matching. *International Conference on Image Processing, ICIP*, pages 517–520, 2003.

[6] M. Diligenti, M. Gori, M. Maggini, and E. Martinelli. Adaptive graphical pattern recognition for the classification of company logos. *Pattern Recognition*, pages 2049–2061, 2001.

[7] D. Doermann, E. Rivlin, and I. Weiss. Applying algebraic and differential invariants for logo recognition. *Machine Vision and Applications*, pages 73–86, 1996.

[8] D. Doermann, E. Rivlin, and I. Weiss. Logo recognition using geometric invariants. *Proceedings of the Second International Conference on Document Analysis and Recognition*, pages 894–897, 2002.

[9] E. Francesconi, P. Frasconi, M. Gori, S. Marinai, J. Q. Sheng, G. Soda, and A. Sperduti. Logo recognition by recursive neural networks. *the Second International Workshop on Graphics Recognition, Algorithms and Systems*, pages 104–117, 1998.

[10] A. Gordo and E. Valveny. A rotation invariant page layout descriptor for document classification and retrieval. *10th International Conference on Document Analysis and Recognition, ICDAR09*, pages 481–485, 2009.

[11] M. Gori, M. Maggini, S. Marinai, J. Q. Sheng, and G. Soda. Edge back propagation for noisy logo recognition. *Pattern Recognition*, pages 103–110, 2003.

[12] A. Hesson and D. Androutsos. Logo and trademark detection in image using wavelet co-occurrence histograms. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, pages 1233–1236, 2008.

[13] J. Luo and D. Crandall. colour object detection using spatial colour joint probability functions. *IEEE Transaction On Image Processing*, pages 1443–1453, 2006.

[14] T. D. Pham. Variogram-based feature extraction for neural-network recognition of logos. *Applications of Artificial Neural Networks in Image Processing*, pages 22–29, 2003.

[15] R. Phan and D. Androutsos. Logo and trademark retrieval in general images database using colour edge gradient co-occurrence histograms. *4th International Conference on Image Analysis and Recognition, ICIAR*, pages 674–685, 2007.

[16] R. Phan and D. Androutsos. Content-based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms. *Computer Vision and Image Understanding*, pages 66–84, 2009.

[17] R. Phan, J. Chia, and D. Androutsos. Colour logo and trademark detection in unconstrained images using colour edge gradient co-occurrence histograms. *Canadian Conference on Electrical and Computer Engineering, CCECE*, pages 531–534, 2008.

[18] A. Wheeler. *Designing Brand Identity.* John Wiley and Sons, 2006.

[19] L. Xia, F. Qi, and Q. Zhou. A learning based logo recognition algorithm using sift and efficient correspondence matching. *International Conference on Information and Automation, ICIA*, pages 1767–1772, 2008.

[20] G. Zhu and D. Doermann. Automatic document logo detection. *Ninth International Conference on Document Analysis and Recognition, ICDAR*, pages 864–868, 2007.

[21] G. Zhu and D. Doermann. Logo matching for document image retrieval. *10th International Conference on Document Analysis and Recognition, ICDAR*, pages 606–610, 2009.

[22] K. J. Zyga, S. Division, J. Schroeder, and R. Price. Logo recognition using retinal coding. *Conference Record of the Thirty-Eighth Asilomar Conference on Signals Systems and Computers*, pages 1549–1553, 2004.